

[illegible]

```

XX WP1; 1994-058406/08.
DR
XX Use of coconut foliar decay virus DNA as promoter - for
PT tissue-specific gene expression in transgenic plants
XX
XX Claim 1; Fig 2; 8pp; German.
PS
CC A DNA fragment from the CPDV genome can be used as a phloem-specific
CC promoter in the construction of transgenic plants. The promoter is
CC strong; it has 30-50% of the activity of the CamV 35S promoter in
CC tobacco plants.
XX
SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match          100.0%; Score 31; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 9,5e-05;
Matches    31; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

OY      1 agccgcgagggaataactagcccccgagct 31
        |||||||
DB       1 agccgcgggggaataactagcccccgagct 31

RESULT      2
AAAX02688
ID AAAX02688 standard; DNA; 31 BP.
XX
AC AAAX02688;
XX
DT 10-MAY-1999 (first entry)
XX
DE Gemini virus DNA fragment stem loop.
XX
DE Stem loop; coconut foliar decay virus; CPDV; bacterial promoter;
KM yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.
XX
OS Gemini virus.
XX
XX Key Location/Qualifiers
FH 1..31
FT stem_loop /*tag= a
FT misc_binding 1..11
FT /*tag= b
FT misc_binding /note= "Region binds to nucleotides 21 to 31"
FT 21..31
FT /*tag= c
FT /*note= "Region binds to nucleotides 1 to 11"
XX
XX DE19730502-A1.
XX PD 21-JAN-1999.
XX PF 16-JUL-1997; 97DE-1030502.
XX PR 16-JUL-1997; 97DE-1030502.
XX
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX
DR WP1; 1999-096863/09.
XX PT Coconut foliar decay virus promoters - for gene expression in
XX bacteria and yeasts
XX
XX Disclosure; Fig 2; 14pp; German.
XX
CC This invention describes a coconut foliar decay virus (CPDV) DNA fragment
CC that includes the stem-loop structure of CPDV DNA but lacks the
CC translation start codons of open reading frames ORF1 and/or ORF2. The
CC new CPDV DNA fragment is useful as a bacterial or yeast promoter, as a

```

```
CC promoter for tissue-specific (especially phloem-specific) gene expression
CC in plants and for production of chimeric constructs for transient or
CC stable expression. Certain fragments of CPDV DNA have stronger promoter
CC activity in E. coli than the CaMV 35S promoter.
XX
SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match          100.0%; Score 31; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 9.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 acccgcgagggaataactacagccccgcgagct 31
         |||
Db       1 agccgcgggggtatatactagccccgcgagct 31

RESULT      3
AAK02686
ID   AAK02686 standard; DNA; 31 BP.

XX      AAX02686;
XX
DT      10-MAY-1999 (first entry)
XX
DE      Gemini virus DNA fragment stem loop.
XX
KW      Stem loop; coconut foliar decay virus; CPDV; bacterial promoter;
XX      yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
OS      Gemini virus.
XX
FH      Key Location/Qualifiers
FT      stem_loop 1..31 /*tag= a
FT      misc_binding 1..11 /*tag= b
FT      misc_binding /*note= "Region binds to nucleotides 21 to 31"
FT      misc_binding 21..31 /*tag= c
FT      FT /*note= "Region binds to nucleotides 1 to 11"
XX
PN      DE19730535-A1.
XX
XX      21-JAN-1999.
XX
PF      16-JUL-1997; 97DE-1030535.
XX
PR      16-JUL-1997; 97DE-1030535.
XX
(PLOC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA      Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
PI      WPI; 1999-096867/09.
DR
XX
PT      Coconut foliar decay virus promoters - for gene expression in
PS      plants, bacteria and yeasts
XX
XX      Disclosure; Fig 2; 12pp; German.
XX
XX      This invention describes a coconut foliar decay virus (CPDV) DNA fragment
XX      that includes the stem-loop structure of CPDV DNA but lacks the
XX      translation start codons of open reading frames ORF1 and/or ORF2. The
XX      new CPDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX      promoter for tissue-specific (especially phloem-specific) gene expression
XX      in plants and for production of chimeric constructs for transient or
XX      stable expression. Certain fragments of CPDV DNA have stronger promoter
XX      activity in E. coli than the CaMV 35S promoter.
XX
SQ      Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;
```

[illegible]

PT plants, bacteria and yeasts
 XX
 PS Disclosure; Fig 2; 12pp; German.
 CC
 CC This invention describes a coconut foliar decay virus (CFDV) DNA fragment
 CC that includes the stem-loop structure of CFDV DNA but lacks the
 CC translation start codons of open reading frames ORF1 and/or ORF2. The
 CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
 CC promoter for tissue-specific (especially phloem-specific) gene expression
 CC in plants and for production of chimeric constructs for transient or
 CC stable expression. Certain fragments of CFDV DNA have stronger promoter
 CC activity in E. coli than the CaMV 35S promoter.
 XX
 SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 63.9%; Score 19.8; DB 20; Length 31;
 Best Local Similarity 77.4%; Pred. No. 11;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agccgcgggggtaactagccccgcgcgt 31
 |||||
 Db 31 AGCCGCGGGGCTAGTATTACCCCGGGGCT 1

RESULT 9
 ID AAS68730/c
 XX AAS68730 standard; cDNA; 686 BP.
 AC AAS68730;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #4534.
 XX
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG04543.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 1; SPO ID No 4534; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 686 BP; 143 A; 236 C; 191 G; 116 T; 0 other;

Query Match 61.3%; Score 19; DB 23; Length 686;
 Best Local Similarity 81.5%; Pred. No. 33;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 ggcgggggtaactagccccgcgcgt 31
 |||||
 Db 251 GCGGGGGTGTACTCTCCCGGGGCT 225

RESULT 10
 ID AAV27112
 XX AAV27112 standard; DNA; 35099 BP.
 AC AAV27112;
 XX
 DT 28-SEP-1998 (first entry)
 XX
 DE Adenovirus 17.
 XX
 KM Adeno virus 17; Ad17; vector; gene therapy; cystic fibrosis;
 KM alpha-antitrypsin deficiency; respiratory disease; ss.
 XX
 OS Mastadenovirus 17.
 XX
 PN WO9822609-A1.
 PD 28-MAY-1998.
 XX
 PF 20-NOV-1997; 97WO-US21494.
 XX
 PR 20-NOV-1996; 96US-0752760.
 XX
 PA (GENZ) GENZYME CORP.
 XX
 PI Armentano DE, Gregory RJ, Smith AE;
 XX
 DR WPI: 1998-312493/27.
 XX
 PT Recombinant adenovirus type 2 vector - useful for targeting
 PT biologically active proteins, used to, e.g. treat cystic fibrosis
 XX
 PS Disclosure; Page 28-38; 67pp; English.
 XX
 CC This is the complete nucleotide (nt) sequence of adenovirus 17
 CC (Ad17). A claimed chimeric adenoviral vector comprises a nt
 CC sequence of an adenovirus in which all or part of a gene encoding
 CC a protein facilitating attachment to, or internalisation into, a
 CC target mammalian cell is replaced by all or part of the
 CC corresponding gene for a second adenovirus from subgroup D,
 CC preferably selected from Ad9, Ad15, Ad17, Ad19, Ad20, Ad22, Ad26,
 CC Ad27, Ad28, Ad30 and Ad39. The vector further comprises a
 CC transgene operably linked to a eukaryotic promoter to allow
 CC expression in a mammalian cell. The replaced gene preferably
 CC encodes an Ad fibre such as Ad17 fibre (see AAV27113) and/or an Ad
 CC penton base such as Ad17 penton base (see AAV27114). The chimeric
 CC vector is used to target biologically active proteins to airway
 CC epithelial cells, especially for gene therapy of diseases such as
 CC cystic fibrosis or alpha-antitrypsin deficiency. The vectors are

CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 490 BP; 139 A; 107 C; 93 G; 140 T; 11 other;

Query Match

Best Local Similarity 59.4%; Score 18.4; DB 22; Length 490;
 Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 3 gccgcggggtaatactagcccccgcgcg 31
 |||||
 DB 414 ccgcgggggaaatacattccctgggggt 442

RESULT 13

ABL19721 standard; DNA; 3894 BP.

ABL19721;

26-MAR-2002 (first entry)

Drosophila melanogaster genomic polynucleotide SEQ ID NO 10636.

Drosophila; developmental biology; cell signalling; insecticide;

pharmaceutical; gene; ds.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -

Claim 1; SEQ ID NO 10636; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 sequences (ABL101840-ABL16175) and the encoded proteins
 (AAB57737-AB572072).
 The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 3894 BP; 1138 A; 918 C; 920 G; 918 T; 0 other;

Query Match 59.4%; Score 18.4; DB 23; Length 3894;
 Best Local Similarity 78.6%; Pred. No. 72;
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 gccgcggggtaatactagcccccgcg 29
 |||||
 DB 1012 gcgcgggggaaatacattccctgggggt 1039

RESULT 14

ABL15567 standard; cDNA; 4263 BP.

ABL15567;

26-MAR-2002 (first entry)

Drosophila melanogaster expressed polynucleotide SEQ ID NO 41183.

Drosophila; developmental biology; cell signalling; insecticide;

pharmaceutical; gene; ss.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -

Claim 1; SEQ ID NO 41183; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 sequences (ABL101840-ABL16175) and the encoded proteins
 (AAB57737-AB572072).
 The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 4263 BP; 1252 A; 1009 C; 990 G; 1012 T; 0 other;

Query Match 59.4%; Score 18.4; DB 23; Length 4263;
 Best Local Similarity 78.6%; Pred. No. 73;
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 gccgcggggtaatactagcccccgcg 29
 |||||
 DB 1381 gcgcgggggaaatacattccctgggggt 1408

RESULT 15

ABL19720/C
 ABL19720 standard; DNA; 6028 BP.

```

XX ABL19720;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 10633.
DE
XX Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
KW
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
PA
XX (PEKE ) PE CORP NY.
PI
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 10633; 21bp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL1840-ABL16175) and the encoded proteins
XX (ABBS7737-ABBS72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ
Sequence 6028 BP; 1486 A; 1441 C; 1361 G; 1740 T; 0 other;

```

Query Match 59.4%; Score 18.4; DB 23; Length 6028;

Best Local Similarity 78.6%; Pred. No. 75;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 gccgcgggggtatatactagcccccgcgg 29

DB 3883 GCGCGGAGGAGGAAAACTAGCCGCCGAGG 3856

Search completed: June 25, 2002, 14:42:05
Job time: 146 sec

BASE COUNT 138 a 299 c 282 g 177 t 9 others

ORIGIN

/db_xref="taxon:99883"

/clone="183N12"

/note="Genoscope sequence ID : COAG183DG06LPI-end : T7"

Query Match 69.0%; Score 21.4; DB 12; Length 905;

Best Local Similarity 80.6%; Pred. No. 1.7e+02;

Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 agccggggggtactatagcccccgcgct 31

482 AGCCGCGGCTACTACCGCCGCCGCGCT 512

RESULT 2

CNS02GRK 1072 bp DNA linear GSS 13-MAY-2000

LOCUS Tetradon nigroviridis genome survey sequence PUC-ori end of clone

DEFINITION 138009 of library G from Tetradon nigroviridis, genomic survey

ACCESSION AL196697.1 GI:7834847

VERSION GSS: genome survey sequence.

KEYWORDS Tetradon nigroviridis.

SOURCE Tetradon nigroviridis.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetradon.

REFERENCE 1 (bases 1 to 1072)

AUTHORS Roest-Crolius H., Jallion O., Dasilva C., Fizames C., Fisher C., Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and Weissenbach J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1072)

AUTHORS Roest-Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C., Bernot A., Fizames C., Wincker P., Brothier P., Quetier F., Saurin W. and Weissenbach J.

TITLE Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 1072)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at

FEATURES

source

1.1072

/db_xref="taxon:99883"

/clone="138009"

/note="Genoscope sequence ID : COAG138AH05SPL-end : PUC-ori"

BASE COUNT 210 a 309 c 319 g 217 t 17 others

ORIGIN

Query Match 69.0%; Score 21.4; DB 12; Length 1072;

Best Local Similarity 80.6%; Pred. No. 1.8e+02;

Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

RESULT 3

AG144021 1360 bp DNA linear GSS 08-JAN-2002

LOCUS Pan troglodytes DNA, clone: RP43-004007.T7, genomic survey

DEFINITION sequence.

ACCESSION AG144021.1 GI:16673699

VERSION GSS: GSS (genome survey sequence).

KEYWORDS Pan troglodytes male lymphocytes DNA, clone.lib.RP43-43 Chimpanzee

SOURCE Male BAC Library clone:RP43-004007.T7.

ORGANISM Pan troglodytes

REFERENCE 1 (sites)

AUTHORS Fujiyama A., Hattori M., Toyoda A., Taylor T.D., Yada T., Toki Y., Watanabe H. and Sakaki Y.

TITLE BAC end sequences of library RP43-43

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1360)

AUTHORS Fujiyama A., Hattori M., Toyoda A., Taylor T.D., Yada T., Toki Y., Watanabe H. and Sakaki Y.

TITLE Direct Submission

COMMENT Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpanzee@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170) Clones are derived from the chimpanzee BAC library RP43-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: T7

LIBRARY

Vector : pBACe3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI

Location/Qualifiers

1.1360

/db_xref="taxon:9598"

/clone="RP43-004007.T7"

/sex="male"

/cell_type="lymphocytes"

/clone.lib="RP43-43 Chimpanzee Male BAC Library"

BASE COUNT 319 a 413 c 306 g 291 t 31 others

ORIGIN

Query Match 66.5%; Score 20.6; DB 12; Length 1360;

Best Local Similarity 85.2%; Pred. No. 3.7e+02;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 ccgcgggggtactatagcccccgcg 29

DB 62 CCGCGGCTGTACTTGCCTCCGCGG 88

RESULT 4

BE968958 876 bp mRNA linear EST 04-OCT-2000

LOCUS 601649921R1 NIH-MGC_74 Homo sapiens cDNA clone IMAGE:3933773.3,

DEFINITION mRNA sequence.

ACCESSION BE968958

VERSION BE968958

KEYWORDS EST.

SOURCE human.

CM892 row: 1 column: 06
ity sequence start: 48
ity sequence stop: 66.
Location/Qualifiers

204 a 225 c 287 g 149 f

AW731193	AW731193	1287 bp	mrna	linear	EST 08-MAR-200
LOCUS					
DEFINITION	Gossypium arboreum 7-10 dpa fiber library Gossypium				

201 Kildare Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktung@uiastate.edu

FEATURES	Location/Qualifiers
source	1. .557

BASE COUNT	146 a	170 c	137 g	104 t
ORIGIN				

```

QY      2  gccgcgggggtaatactagccccgcgcgc 30
          | | | | | | | | | | | | | |
Db      389 GACGCGGTGTAAGTATTAGTCCCGCGGC 36

```

SOURCE	human.
ORGANISM	Homo sapiens

JOURNAL
COMMENT

FEATURES
source

```
/clone_id="HTB"  
/tissue_type="Hypothalamus"
```

BASE COUNT	185 a	124 c	129 g	190 t	1 others
ORIGIN					

Best Local Similarity 79.3%; Pred No. 9.9e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0;
Gaps 0;

Db 597 GCCGCTGGCGTAAATAGCAGCCCGCGCTGC 623

SOURCE	domestic silkworm
ORGANISM	<i>Bombyx mori</i>

JOURNAL Unpublished (1999)
COMMENT Contact: Mita K

Email: kmita@nirs.go.jp
PROJECT = 'CREST project by JST'.

BASE COUNT	173 a	183 c	191 g	166 t	1 others
ORIGIN					

Db 398 AGCCGCGTGTAAGTCCGCTCGGG 370

DEFINITION
6024853295b6f1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4593646 5',
mRNA sequence.
BCA01207

REFERENCE
1 (bases 1 to 797)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."

BASE COUNT 256 a 243 c 230 g 110 t
ORIGIN

Query Match 62.6%; Score 19.4; DB 10; Length 839;
Best Local Similarity 79.3%; Pred. No. 1e+03;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 agccgcgggggtactactagccccgcgg 29
||||| ||||| ||||| ||||| |||||
Db 766 AGCCAGGGGGGAGACTACGAGACCCCGCGG 794

RESULT 15

BF300720 897 bp mRNA linear EST 21-NOV-2000
LOCUS 602031903F1 NCI_CGAP_SG2 Mus musculus cDNA IMAGE:4166784 5',
DEFINITION mRNA sequence.

ACCESSION BF300720
VERSION BF300720.1 GI:11247243
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE

1 (bases 1 to 897)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:

<http://image.lnl.gov>
Plate: L1AM9456 row: F column: 01
High quality sequence stop: 571.

FEATURES
source location/Qualifiers
1..897

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:4166784"

/clone_id="NCI_CGAP_SG2"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1;
MolI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

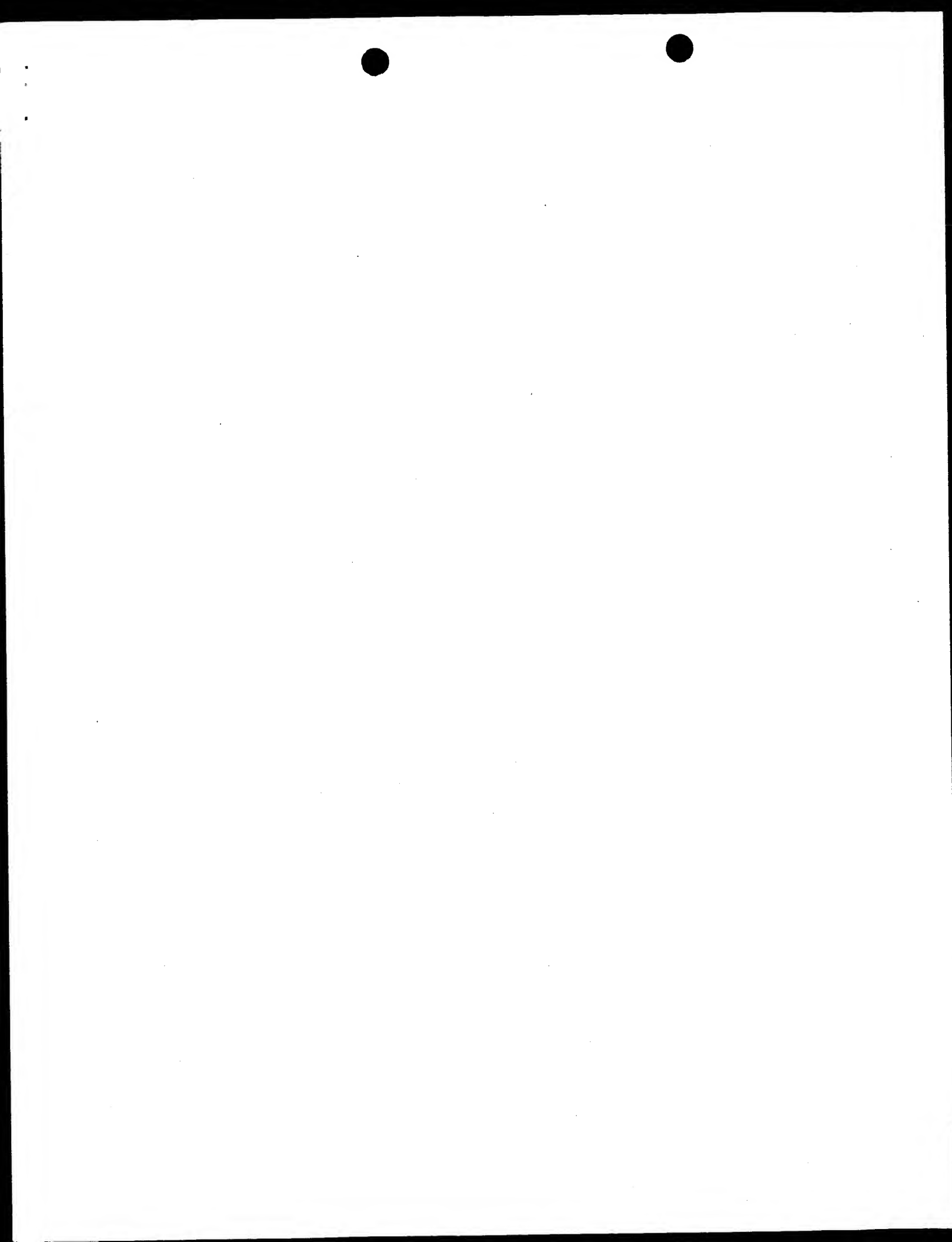
BASE COUNT 266 a 216 c 234 g 180 t
ORIGIN 1 others

Query Match 62.6%; Score 19.4; DB 10; Length 897;
Best Local Similarity 79.3%; Pred. No. 1e+03;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 gccgcgggggtactactagccccgcggc 30
||||| ||||| ||||| ||||| |||||
Db 660 GCCGCGGGGAGAACACACGCCCGCGCGC 688

Search completed: June 25, 2002, 15:09:00
Job time: 1761 sec



TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-752-760A-1

Query Match 60.6%; Score 18.8; DB 2; Length 35081;
Best Local Similarity 76.7%; Pred. No. 22;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagcccccgcgcgc 30
DB 15096 ACCGCTGGGGTATTACTAGCCCGACGACG 15125

RESULT 6
US-08-973-068-15/c
Sequence 15, Application US/08973068
Patent No. 6127604
GENERAL INFORMATION:
APPLICANT: Dale, James Langham
APPLICANT: Harding, Robert Maxwell
APPLICANT: Dugdale, Benjamin
APPLICANT: Beetham, Peter Ronald
APPLICANT: Hafner, Gregory John
TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
FILE REFERENCE: 09657/002001
CURRENT APPLICATION NUMBER: US/08/973,068
CURRENT FILING DATE: 1998-03-12
EARLIER APPLICATION NUMBER: PCT/AU96/00335
EARLIER FILING DATE: 1996-05-31
NUMBER OF SEQ ID NOS: 61
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15
LENGTH: 60
TYPE: DNA
ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-973-068-15

Query Match 58.7%; Score 18.2; DB 3; Length 60;
Best Local Similarity 74.2%; Pred. No. 25;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagcccccgcgcgc 31
DB 47 AGCACGGGGGTAAATAGTCCCGCGCT 17

RESULT 7
US-08-973-068-12/c
Sequence 12, Application US/08973068
Patent No. 6127604
GENERAL INFORMATION:
APPLICANT: Dale, James Langham
APPLICANT: Harding, Robert Maxwell
APPLICANT: Dugdale, Benjamin
APPLICANT: Beetham, Peter Ronald
APPLICANT: Hafner, Gregory John
APPLICANT: Becker, Douglas Kenneth
TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
FILE REFERENCE: 09657/002001
CURRENT APPLICATION NUMBER: US/08/973,068
CURRENT FILING DATE: 1998-03-12
EARLIER APPLICATION NUMBER: PCT/AU96/00335
EARLIER FILING DATE: 1996-05-31
NUMBER OF SEQ ID NOS: 61
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 69
TYPE: DNA
ORGANISM: Banana Bunchy Top Virus (BBTV)

US-08-973-068-12

Query Match 58.7%; Score 18.2; DB 3; Length 69;
Best Local Similarity 74.2%; Pred. No. 25;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagcccccgcgcgc 31
DB 56 AGCGCTGGGGTAAATAGTCCCGACGCGCT 26

RESULT 8
US-08-973-068-30/c
Sequence 30, Application US/08973068
Patent No. 6127604
GENERAL INFORMATION:
APPLICANT: Dale, James Langham
APPLICANT: Harding, Robert Maxwell
APPLICANT: Dugdale, Benjamin
APPLICANT: Beetham, Peter Ronald
APPLICANT: Hafner, Gregory John
APPLICANT: Becker, Douglas Kenneth
TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
FILE REFERENCE: 09657/002001
CURRENT APPLICATION NUMBER: US/08/973,068
CURRENT FILING DATE: 1998-03-12
EARLIER APPLICATION NUMBER: PCT/AU96/00335
EARLIER FILING DATE: 1996-05-31
NUMBER OF SEQ ID NOS: 61
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 352
TYPE: DNA
ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-973-068-30

Query Match 58.7%; Score 18.2; DB 3; Length 352;
Best Local Similarity 74.2%; Pred. No. 28;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagcccccgcgcgc 31
DB 104 AGCACGGGGGTAAATAGTCCCGCGCT 74

RESULT 9
US-08-973-068-24/c
Sequence 24, Application US/08973068
Patent No. 6127604
GENERAL INFORMATION:
APPLICANT: Dale, James Langham
APPLICANT: Harding, Robert Maxwell
APPLICANT: Dugdale, Benjamin
APPLICANT: Beetham, Peter Ronald
APPLICANT: Hafner, Gregory John
APPLICANT: Becker, Douglas Kenneth
TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
FILE REFERENCE: 09657/002001
CURRENT APPLICATION NUMBER: US/08/973,068
CURRENT FILING DATE: 1998-03-12
EARLIER APPLICATION NUMBER: PCT/AU96/00335
EARLIER FILING DATE: 1996-05-31
NUMBER OF SEQ ID NOS: 61
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 24
LENGTH: 547
TYPE: DNA
ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-973-068-24

Query Match 58.7%; Score 18.2; DB 3; Length 547;
Best Local Similarity 74.2%; Pred. No. 29;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtactactagccccgcgct 31
DB 366 AGCGCTGGGGGTATATAGTCCCGCGCT 336

RESULT 10
US-08-973-068-29/c

; Sequence 29, Application US/08973068
; Patent No. 6127604

; GENERAL INFORMATION:

; APPLICANT: Dale, James Langham

; APPLICANT: Harding, Robert Maxwell

; APPLICANT: Dugdale, Benjamin

; APPLICANT: Beetham, Peter Ronald

; APPLICANT: Hainer, Gregory John

; APPLICANT: Becker, Douglas Kenneth

; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS

; FILE REFERENCE: 09657/002001

; CURRENT APPLICATION NUMBER: US/08/973,068

; EARLIER FILING DATE: 1998-03-12

; EARLIER APPLICATION NUMBER: PCT/AU96/00335

; EARLIER FILING DATE: 1996-05-31

; NUMBER OF SEQ ID NOS: 61

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 29

; LENGTH: 622

; TYPE: DNA

; ORGANISM: Banana Bunchy Top Virus (BBTV)

US-08-973-068-29

Query Match 58.7%; Score 18.2; DB 3; Length 622;
Best Local Similarity 74.2%; Pred. No. 29;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtactactagccccgcgct 31
DB 374 AGCAGGGGGGTATATAGTCCCGCTGCT 344

RESULT 11

US-08-973-068-27/c

; Sequence 27, Application US/08973068

; Patent No. 6127604

; GENERAL INFORMATION:

; APPLICANT: Dale, James Langham

; APPLICANT: Harding, Robert Maxwell

; APPLICANT: Dugdale, Benjamin

; APPLICANT: Beetham, Peter Ronald

; APPLICANT: Hainer, Gregory John

; APPLICANT: Becker, Douglas Kenneth

; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS

; FILE REFERENCE: 09657/002001

; CURRENT APPLICATION NUMBER: US/08/973,068

; EARLIER FILING DATE: 1998-03-12

; EARLIER APPLICATION NUMBER: PCT/AU96/00335

; EARLIER FILING DATE: 1996-05-31

; NUMBER OF SEQ ID NOS: 61

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 27

; LENGTH: 624

; TYPE: DNA

; ORGANISM: Banana Bunchy Top Virus (BBTV)

US-08-973-068-27

Query Match 58.7%; Score 18.2; DB 3; Length 624;
Best Local Similarity 74.2%; Pred. No. 29;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtactactagccccgcgct 31
DB 375 AGCAGGGGGGTATATAGTCCCGCTGCT 345

RESULT 12
US-08-202-186-20/c

; Sequence 20, Application US/08202186

; Patent No. 5756708

; GENERAL INFORMATION:

; APPLICANT: KARAN, Mirko

; APPLICANT: BURNS, Thomas M.

; APPLICANT: DALE, James L.

; APPLICANT: HARDING, Robert M.

; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W.

; CITY: Washington, D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/202,186

; FILING DATE: 24-FEB-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: JEFFERY, Donald D.

; REGISTRATION NUMBER: 19,980

; TELEPHONE: 202 672 5300

; TELEFAX: 202 672 5399

; TELEEX: 904136

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1075 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: circular

US-08-202-186-20

Query Match 58.7%; Score 18.2; DB 1; Length 1075;
Best Local Similarity 74.2%; Pred. No. 30;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtactactagccccgcgct 31
DB 31 AGCGCTGGGGGTATATAGTCCCGCGCT 1

RESULT 13

US-08-973-068-2/c

; Sequence 2, Application US/08973068

; Patent No. 6127604

; GENERAL INFORMATION:

; APPLICANT: Dale, James Langham

; APPLICANT: Harding, Robert Maxwell

; APPLICANT: Dugdale, Benjamin

; APPLICANT: Beetham, Peter Ronald

; APPLICANT: Hainer, Gregory John

; APPLICANT: Becker, Douglas Kenneth

; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS

; FILE REFERENCE: 09657/002001

; CURRENT APPLICATION NUMBER: US/08/973,068

; CURRENT FILING DATE: 1998-03-12

```

1  EARLIER APPLICATION NUMBER: PCT/AU96/003355
2  EARLIER FILING DATE: 1996-05-31
3  NUMBER OF SEQ. ID NOS: 61
4  SOFTWARE: FASTSEQ for Windows Version 3.0
5  SEQ. ID NO. 2
6  LENGTH: 1075
7  TYPE: DNA
8  ORGANISM: Banana Bunchy Top Virus (BBTV)
9  FEATURE:
10 NAME/KEY: CDS
11 LOCATION: (213)...(737)
12 US-08-973-068-2

```

Query Match	58.7%;	Score 18.2;	DB 3;	Length 1075;
Best Local Similarity	74.2%;	Pred. No. 30;		
Matches 23;	Conservative 0;	Mismatches 8;	Indels 0;	Gaps 0;

```

QY      1 agccgcggggtactactagccccgcgcgt 31
        ||| ||||| ||| ||| |||
Db      31 AGCGCTGGGGTAAATAGTCCCAAGCGCT 1

```

RESULT 14
US-08-202-186-23/c
Sequence 23, Application US/08202186
Patent No. 5756708

GENERAL INFORMATION:
APPLICANT: KARAN, Mirko
APPLICANT: BURNS, Thomas M.
APPLICANT: DALE, James L.
APPLICANT: HARDING, Robert M.
TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA

2 ZP: 20007-5109
3 COMPUTER READABLE FORM:
4 MEDIUM TYPE: Floppy disk
5 COMPUTER: IBM PC compatible
6 OPERATING SYSTEM: PC-DOS/MS-DOS
7 SOFTWARE: Patent in Release #1.0, Version #1.25
8 CURRENT APPLICATION DATA:
9 APPLICATION NUMBER: US/08/202,186
10 FILING DATE: 24-FEB-1994

```

: INFORMATION FOR SEQ. ID NO.: 23
:
: SEQUENCE CHARACTERISTICS:
:
:   LENGTH: 1089 base pairs
:   TYPE: nucleic acid
:   STRAIGHTNESS: single
:   TOPOLOGY: circular
:
US-08-202-186-23

```

Query Match	58.7%;	Score 18.2;	DB 1;	Length 1089;
Best Local Similarity	74.2%;	Pred. No. 30;		
Matches 23;	Conservative 0;	Mismatches 8;	Indels 0;	Gaps 0

QY 1 agccgcggyggtataactagccccgcgct 31
||| ||||| ||| ||| |||
Db 31 AGCAGCGGCGGTATATAGTCCCCCGTGCT 1

```

RESULT 15
US-08-973-068-8/C
: Sequence 8, Application US/08973068
: Patent No. 6127604
: GENERAL INFORMATION:
: APPLICANT: Dale, James Langham
: APPLICANT: Harding, Robert Maxwell
: APPLICANT: Dugdale, Benjamin
: APPLICANT: Beetham, Peter Ronald
: APPLICANT: Hainer, Gregory John
: APPLICANT: Becker, Douglas Kenneth
: TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
: FILE REFERENCE: 09657/002001
: CURRENT APPLICATION NUMBER: US/08/973,068
: CURRENT FILING DATE: 1998-03-12
: EARLIER APPLICATION NUMBER: PCT/A096/00335
: EARLIER FILING DATE: 1996-05-31
: NUMBER OF SEQ ID NOS: 61
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 8
: LENGTH: 1089
: TYPE: DNA
: ORGANISM: Banana Bunchy Top Virus (BBTV)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (281)...(742)
US-08-973-068-8-

```

Query March	58.7%	Score 18.2;	DB 3;	Length 1089;
Best Local Similarity	74.2%	Pred. No. 30;		
Matches 22;	Conservative	0;	Mismatches	8; Indels 0; Gaps
QY	1	aagccggcggygtataactagccccgcgagct	31	
Dd	31	AeCAGCGGGGGTAATATAgTCCCCGCTCT	1	

Search completed: June 25, 2002, 14:40:33
Job time: 54 sec

Search completed: June 25, 2002, 14:40:33
Job time: 54 sec

|||||
 CCGCGGCT
 30

```

TATGTGATCATTAATTACACCAATTACCCGCGCCGACCGCTATGCTTTACATCTTANGAATATTCCTGCC
590      600      610      620      630      640      650
720      730      740      750      760      770      780
CAGTTTATTTCTGTGCTCCCTGCTGATTTTCAATATATACGACGACGACATTTGCCAATATACAGATGACATG
|||      |||      |||      |||      |||      |||      |||
CAGGCGCAAGGCGGAGGCTCT-----ACCGGCGGAAGCGCGGAACAAATATGATGAGTTATG
660      670      680      690      700      710      720
790      800      810      820      830      840      850
GTCCAGACCCCAAGAT-AACTAAGGGGCTTCGTATTTGTCCGAACCTGAATCCCTGTCTTACACATCTTACACA
|||      |||      |||      |||      |||      |||      |||
GGCGGCGCCCAATAAAGATTCATTGTGATAGAACGAATCTGTACTTCTGTTCAGTGCACGCAACCA
730      740      750      760      770      780      790
860      870      880      890      900      910      920
GGGATATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA
|||      |||      |||      |||      |||      |||      |||
CTTCCAGCTACCAATCCAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT
800      810      820      830      840      850      860
930      940      950      960      970      980      990
TACGTGTACATTAATTCCTCTTCTGTTCCACACATGTGTAGAAC--AGTCGGGCTTGATCCGAGATAT
|||      |||      |||      |||      |||      |||      |||
TGTTTTCGGGATCGGACGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT
870      880      890      900      910      920      930
1000     1010     1020     1030     1040     1050     1060     1070
TTGGCAAAACAGCGACTTCCGCTCTCTCTCCGATATCCAGAGATTTGTCGATCGTCCGCTGCG
|||      |||      |||      |||      |||      |||      |||
ATACCGCGGGGGCTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT
940     950     960     970     980     990     1000     1010
1080     1090     1100     1110     1120     1130     1140     1150
TCTCCGATTCGACAGACACTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAG
|||      |||      |||      |||      |||      |||      |||
CCTCCATTCCGCGGCTGCTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT
1020     1030     1040     1050     1060     1070     1080
1150     1160     1170     1180     1190     1200
TCTGTGTCAT-----TCCACCGAGCTCCGCTGTCACAGCATCTTCGCTATCCGCTGCGCTCTTCAGG
|||      |||      |||      |||      |||      |||      |||
TCGAGTCCCTTAATCTGTATCTATAGTTGGGAGGTTGACCTTCAACTGCGCAAGACACACCTCC
1090     1100     1110     1120     1130     1140     1150
1210     1220     1230     1240     1250     1260     1270
CGGAGTTTCATCAGCTTCTCAGCAAAATCGTTGGGCAATGCTGCTTTGACCTCCAGAGAGAGTGGGACT
|||      |||      |||      |||      |||      |||      |||
AAGATTTCATCAGCTTGAAGACCGGCTGGGAGCTGCAAGATTGAAGACTGTTCTTGGAAATGACAGGATTC
1160     1170     1180     1190     1200     1210     1220
1280
X
CCGTGCTCGAG
|||
ACCTGAGAGCGGACCGGCTGCTCCGACGACAGATAGACTACTGTTGAGAGAACGGGT
1230     1240     1250     1260     1270     1280
270      280      290      300      310      320      330
GCGGGAATGAGAGACCTTTAGCAAGCGGGTTAGCAGAGTTGGGAGCGCGGGGTATATATACCC
|||||      ||||||      ||||||      ||||||      ||||||      ||||||
AGCGCGGGGGTATATACCTAGCC
X
340      350      360      370      380      390      400
CCCCGGGCTATTACTTACTTACGACGGAGCTTAGCAGAGTTTGGCGCCAGATCAACT
X

```

2. US-09-462-955B-1-COPY' (1-1288)
 US-09-462-955B-2 Sequence 2, Application US/09462955B

Initial Score = 31 Optimized Score = 31 Significance = 0.28
 Residue Identity = 100% Matches = 31 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

080 670 680 690 700 710 720
 AATATCTGCCCCAGGCGGAAAGGCGTGGGAGGTGCTACCCGGCCGAAGGCGCGGGAACAATATGATGAGTTA
 |||||

Query Match	100.0%;	Score 31;	DB 6;	Length 31
Best Local Similarity	100.0%;	Pred. No. 0.36;		

Matches 31: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agccggggggaataactagccccggct 31
 1 agccggggggaataactagccccggct 31

RESULT 2
 AR172770/c 1291 bp DNA linear PAT 17-DEC-2001
 LOCUS AR172770 Sequence 1 from patent US 6303345.
 DEFINITION AR172770
 ACCESSION AR172770
 VERSION AR172770.1 GI:17912261
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1291)
 AUTHORS Rohde,W., Becker,D., Randles,J.W., Hehn,A. and Salamin,I.F.
 TITLE Use of a virus DNA as promoter
 JOURNAL Patent: US 6303345-A 1 16-OCT-2001;
 FEATURES Location/Qualifiers
 source 1..1291
 /organism="unknown"

BASE COUNT 336 a 323 c 332 g 300 t
 ORIGIN

Query Match 100.0%; Score 31; DB 6; Length 1291;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agccggggggaataactagccccggct 31
 1 agccggggggaataactagccccggct 40

RESULT 3
 CPDGC 1291 bp ss-DNA circular VRL 02-AUG-1993
 LOCUS Cocoon foliar decay virus, complete genome.
 DEFINITION M29963
 ACCESSION M29963.1 GI:323306
 VERSION
 KEYWORDS circular; complete genome.
 JOURNAL Cocoon foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
 MEDLINE Cocoon foliar decay virus
 COMMENT Cocoon foliar decay virus
 DRAFT entry and printed sequence for [1] kindly submitted by
 W. Rohde, 15-MAR-1989, for release after publication.
 FEATURES
 source 1..1291
 Location/Qualifiers
 1..1291
 /organism="Cocoon foliar decay virus"
 /db_xref="taxon:12474"
 40..70
 /note="stem-loop structure"
 103..975
 /note="ORF 1"
 /codon_start=1
 /protein_id="AA42894.1"
 /db_xref="GI:323307"
 /translation="MGSSIRRCWFTLNTVEEPAANVRIRIESLNLVVAIVGDEVAPS
 TGOHRLQGFHLTKTGRRLGSLKTVLGNDRHLFPTGSDONRDCSKRVLLEHGV
 TRGVKRPRLAQRFAEPDELRLDEPGTRCYVHGASVEMTMAADNPPPPYHMO
 LEVLSAIGEPADORTILWICRGDGGKSVFAVYLGLKPMWFTGGRDVLVITTE
 DPKRNLIDVPCNLLEYLINVALLCYKNRAFSDDKYEPLSYLGFHVAHVLFANVLDP
 VIKSRDRIKLMNT"
 314..775

CDS

/note="ORF 2"
 /codon_start=1
 /protein_id="AA42895.1"
 /db_xref="GI:323308"
 /translation="MTGFTWSPRPVPTNRITETVRRNGCFSTESRLVLEKGDHWP
 DILRLMNSAMKTAQVFDALYTELWNSODGPLKIRFHTITSLKCCLRSEQR
 IASGSGVADEETEGSPCLPNISDSSPTGTHVPEREYCTSTRTONEI"
 complement(422..568)
 /note="ORF 6"
 /codon_start=1
 /protein_id="AA42896.1"
 /db_xref="GI:323309"
 /translation="MEMGTFDFORPILSIPPKLRVQRIRGILGPGVHQPQOLVGP
 IYAF"
 639..797
 /note="ORF 3"
 /codon_start=1
 /protein_id="AA42897.1"
 /db_xref="GI:323310"
 /translation="MRTRRRRREVRCQISRTQARLVLMHMMNKGRIYVHRGPKTK
 FNRCTQV"
 complement(823..987)
 /note="ORF 5"
 /codon_start=1
 /protein_id="AA42898.1"
 /db_xref="GI:323311"
 /translation="MTHTLNIPOFYSVPADFOILRODICKREYMHMVEPKTKGFVY
 RTGCPVLNTE"
 1098..1286
 /note="ORF 4"
 /codon_start=1
 /protein_id="AA42899.1"
 /db_xref="GI:323312"
 /translation="MNRVMGPGPTIKSIWIRNLLLOCTOPLSTPIQVSSLEKK
 AASLYPSICFCAIGRLS"

BASE COUNT 336 a 323 c 332 g 300 t
 ORIGIN

Query Match 100.0%; Score 31; DB 14; Length 1291;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agccggggggaataactagccccggct 31
 1 agccggggggaataactagccccggct 40

RESULT 4
 AE006200/c 9294 bp DNA linear BCF 08-MAR-2001
 LOCUS Pasteurella multocida PM70 section 167 of 204 of the complete
 DEFINITION genome.
 ACCESSION AE006200 AE004439
 VERSION AE006200.1 GI:12722026
 KEYWORDS
 SOURCE Pasteurella multocida.
 ORGANISM Pasteurella multocida.
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 Pasteurella multocida.
 Pasteruella.
 1 (bases 1 to 9294)
 May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
 Kapur,V.
 Complete genomic sequence of Pasteurella multocida, PM70
 Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
 MEDLINE 21145866
 /translation="MGSSIRRCWFTLNTVEEPAANVRIRIESLNLVVAIVGDEVAPS
 TGOHRLQGFHLTKTGRRLGSLKTVLGNDRHLFPTGSDONRDCSKRVLLEHGV
 TRGVKRPRLAQRFAEPDELRLDEPGTRCYVHGASVEMTMAADNPPPPYHMO
 LEVLSAIGEPADORTILWICRGDGGKSVFAVYLGLKPMWFTGGRDVLVITTE
 DPKRNLIDVPCNLLEYLINVALLCYKNRAFSDDKYEPLSYLGFHVAHVLFANVLDP
 VIKSRDRIKLMNT"
 314..775
 location/Qualifiers

source 1. .9294
/organism="Pasteurella multocida"
/strain="PM70"
/db_xref="taxon:747"
complement(161. .853)
gene /gene="PM1626"
complement(161. .853)
CDS /gene="PM1626"
complement(161. .853)
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03710.1"
/db_xref="GI:12722027"
/translation="MKTGILIGMSPESTVYRYRLINOLNQAQSGNSAPLVYVNE
FQIVLOKQGEHKAAGETIATAKKTQIGAGGILLATNMHKAPOIIVIDVPEL
HIDATAKAIRKINLTQVALLGTOFTODDPYRQDLLHGGITPLVPNQKQOETHRII
FELICVQIKRPSKQFQVLDVINDAKQAGVILGCTEIGILLINQDNLTFPFDTTL
HQAQAEIFIRE"
gene complement(872. .1102)
/gene="11VM"
/note="PM1627"
complement(872. .1102)
CDS /gene="11VM"
/codon_start=1
/transl_table=11
/product="11VM"
/protein_id="AAK03711.1"
/db_xref="GI:12722028"
/translation="MQYQFLDVNQRPETLERILRIHRGFYVVAAMNTAQEQAN
IDFYKSDRAEALLINOLSKYVDVLAHQVRA"
gene complement(1140. .2789)
/gene="11VG"
/note="PM1628"
complement(1140. .2789)
CDS /gene="11VG"
/codon_start=1
/transl_table=11
/product="11VG"
/protein_id="AAK03712.1"
/db_xref="GI:12722029"
/translation="MNGARLVTESLKHGVTTFVCGYRGAIMPYDALYDAGIEHLIC
RNEQGAAMAAGVARSQGVCIATSGPAAALITGIDALMDSIPVAVITGVVASP
LIGTDAEQEADVLGLSLACTKHSFLVQSIIEELPVIAKAFETIAGSRGPVLIDIPD
IOIAPDTVPTVYSAKPTLASOHLIDALDILAKARPVLYVGGVGMAGVAYALAE
FLAYTQMPSTVSLKIGSVHPNDPYKMGIMHCTKXANVVOECDLILFAGARFDR
VTGKIDFPAHAKYIHDIDMAEIGKLRPVALCGDIVAFNLSTPLNTAEQAH
KOLQKTHFCSTNQGDFTINPLWLLNSLSKRPQSAVITTDVGQHQMSKOHQHYA
PENYTSAGYVTFMFGFLPAALGAKARPDEVIIISGDSIMANNIOELGTFKRGKTFV
KIVLLDNORLQMVQOELFENARFSMTILDNDPFYMLAAAFGIOGERITKGEVEG
AIDRLLSKEAYFLHVCISSNENWPLVPPMACNLDWVEQW"
gene 3149. .3967
/gene="PM1629"
3149. .3967
CDS /gene="PM1629"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03713.1"
/db_xref="GI:12722030"
/translation="MSFFPRRAVYSDLDGTLNLNHNMIIGFTIOTLOQLAAKIDIML
ATGRNHTDILPLIKKVINIEKAVMTTSGARAQDQGMILVRDLPEDIAPIINDLPD
RQRYVSAYQGDWFINIDVQLRKYHODSGFMVEYDQAHGREFEKYFFIGREPO
DLIGLEOHLQNAVADTISITYSEVLCLEIMKNKVSASALEKVLADRDYDLOHCLAFG
DGMNDVOMLSVSGKGYGNADPRLKACPHLEVLIGNNAQESVANYIRTYFDIE"
gene 3983. .4387
/gene="CrcB"
/note="PM1630"
3983. .4387
CDS /gene="CrcB"
/codon_start=1
/transl_table=11
/product="CrcB"
/protein_id="AAK03714.1"
/db_xref="GI:12722031"
/translation="MAGSRVYSVGGQIIRISSGALGALSRLGILLNPLFSFSL
GITLVANVLGCLTIGVFLAFWQYQCAERLLEFVIGLSITTFSSFAVEIENLQ
QKWLAGLMLASGHLGLGCLTFALGVFTWRWQ"
gene complement(4365. .5318)
/gene="P1dAB"
complement(4365. .5318)
CDS /note="PM1631"
complement(4365. .5318)
/gene="P1dAB"
/codon_start=1
/transl_table=11
/product="P1dAB"
/protein_id="AAK03715.1"
/db_xref="GI:12722032"
/translation="MIREPFTHEPALAEHPRAOPPLQDVQCKKGVRLTYRFEVDN
FVQNNVLVNGRAENLKWTELAYDPEQOQGYVLVDRHGQGYSQLLPDHEKGYID
EFRFYADDMAILLAKISALPYENQHLIAGSLALSTYLAAYDQVSAVSAPFY
G1PLKHSRDELLINMLMLGQGSRYVFGKGRYOPADLDNDSCCRRTRMWRNIR
RYPHILGCTPFRVHLGFSAIKKLPTILPRIEIPVILQSEKEKIVEKNLKLTLAL
LPRELTOIQSKHEILFERDLIRSEALKKISQFTANSK"
gene complement(5460. .6569)
/gene="P1dAB"
/note="PM1632"
complement(5460. .6569)
CDS /gene="P1dAB"
/codon_start=1
/transl_table=11
/product="P1dAB"
/protein_id="AAK03716.1"
/db_xref="GI:12722033"
/translation="MONGVFIQWGNVGSYLMDRMOENDFANINPFFTTSOAGNA
PVGCGQAGELKADFDIAELQKLDITVTCOGGYTNEVPKLRAGMKGYWIDASAL
RMEKDAIIVLDPVNOWNVISEGLKNGIKTYVGGNCTVSMMLGGLFEEDLYENVSVA
TYOASGAGKKNMRELLVOMGLEDVSKVDLNDPASIIDIERKYAKRGLPIENF
GAPLAGSLIPIDKIMWEDQTKREKGYAFTNKLDELSPITVDSGLCVIRGALRHS
QATFIRKKDLPLEITEOILLASHNEWKVI PNKDVTLRELTPTKVTGLSVAVGRLR
KLKMPBYLAAPFTVGDQLMGAEPRLRILKOLVA"
gene 6770. .7009
/gene="PM1633"
6770. .7009
CDS /gene="PM1633"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03717.1"
/db_xref="GI:12722034"
/translation="MDLGFYFAPYRDERQKRSYVYIQTPEKLLGASISLRNFS
SDRESKARHTIDINIKOITBEKNIKRSKIELFQ"
gene 7028. .7381
/gene="PM1634"
7028. .7381
CDS /gene="PM1634"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03718.1"
/db_xref="GI:12722035"
/translation="MNPNNKRIILKRSKGLPFLALAEFEDYFGSVKSDSETKELFS
FVAELMHDELKFAIRGKFLGSIIEQDILDFQAMPDHYDKLEFYDIDHLMWITYAP
AGAVWICEDGYEHW"
gene 7583. .7948
/gene="PM1635"
7583. .7948
CDS /gene="PM1635"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03719.1"
/db_xref="GI:12722036"
/translation="MLEKSEAYKRIIVEENGLVLDGLFANFOECYDIDRDKPEIKE
MFLDLMDLMTGELKATKGLFEGSIEEOIDLFQAMPDHYDKLEFYDIDHLMWIT
TYAPAGAVWIMWEDGYEDWT"
gene 8372. .8637

```

CDs      /gene="PM1636"
          /size="8637
          /gene="PM1636"
          /codon_start=1
          /transl_table=11
          /product="unknown"
          /protein_id="AAK03720.1"
          /db_xref="GI:12722037"
          /translation="MLEKSEAQYKRIVENNRGLITLIGMFSFQVCYGDIRDKEIKE
MPELIMTDMKTGTEIKLATKGRKFLGSEIEQIDLPFQAMPDHDDKELEYDINIMWH
VYAPACAVWICEPGEYEW"
          8707..9144
          /gene="PM1637"
          /size="9144
          /gene="PM1637"
          /codon_start=1
          /transl_table=11
          /product="unknown"
          /protein_id="AAK03721.1"
          /db_xref="GI:12722038"
          /translation="MMLYRWSSRSRSPDERYQDRKENVMLKSESAQYKRIVENNRGL
ALDLPFSFQVCYGDIDKEIKEIMPELIMADLMKTGTEIKLATKGRKFLGSEIEQIDYV
FQAMPDHDDKELEYDINIMWHVYAPACAVWIMEDYEW"

BASE COUNT      2772 a      1733 c      1979 g      2810 t

ORIGIN
Query Match      68.4%; Score 21.2; DB 1; Length 9294;
Best Local Similarity 88.5%; Pred. No. 4e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy      5 gcgggggtaataactagcccccgcggc 30
          ||||| ||||| ||||| |||||
Db      676 GCGGGGGAATACVAGCCATCGCGC 651

RESULT 5
LOCUS      AB000926/c
DEFINITION      Milk vetch dwarf virus genome segment 7 encoding viral
non-structural protein, complete sequence.
ACCESSION      AB000926
VERSION      AB000926.1 GI:3808185
KEYWORDS      viral non-structural protein.
SOURCE      milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
ORGANISM      Pisum sativum
REFERENCE      1 (bases 1 to 981)
AUTHORS      Sano,Y.
TITLE      Direct Submission
JOURNAL      Submitted (08-FEB-1997) Yoshitaka Sano, Kyoto Institute of
Technology, Department of Applied Biology; Matsugasaki, Sakyo-ku,
Kyoto 606-8585, Japan (Tel:075-724-7764, Fax:075-724-7764)
2 (sites)
Sano,Y., Wada,M., Hashimoto,Y., Matsumoto,T. and Kojima,M.
Sequences of ten circular ssDNA components associated with the milk
vetch dwarf virus genome
J Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
99094638
On Oct 29, 1998 this sequence version replaced gi:3798663.
FEATURES
location/Qualifiers
1..981
/organism="milk vetch dwarf virus"
/vision
/db_xref="taxon:67585"
/chromosome="segment 7"
/lab_host="Pisum sativum"
6..34
280..286
346..723
/function="unknown"
/note="putative"
/codon_start=1

```

```

/prefix="viral non-structural protein"
/protein_id="BAA33986.1"
/db_xref="GI:3798664"
/translation="MEDFOPKPSIYGEIYQMEEDAPFSCYHEPIRREYDILGEMCR
RHRKRLPAVPEKLPYAPIRWVLTKAIYVRVDECKSCSHEELSRDYNPIKKEGLKD
LVDGNYRYQVYSSSCNRDKSD"
L730..735
polyA_signal
BASE COUNT      291 a      170 c      214 g      306 t
ORIGIN

Query Match      66.5%; Score 20.6; DB 14; Length 981;
Best Local Similarity 85.2%; Pred. No. 1.le+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      5 ggggggtaataactagcccccggcgact 31
|||||
Db      31 GCGGGGGTAATATCAAGCCCCGCCCT 5

RESULT      6
AB000924/c      989 bp      DNA      circular VRL 30-OCT-1998
LOCUS
DEFINITION
AB000924      989 bp      DNA      circular VRL 30-OCT-1998
non-structural protein, complete sequence.
ACCESSION
AB000924      1 GI:3808183
VERSION
AB000924.1 GI:3808183
KEYWORDS
viral non-structural protein.
SOURCE
milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
milk vetch dwarf virus
VIRUSES: ssDNA viruses; Nanovirus.
1 (bases 1 to 989)
REFERENCE
AUTHORS
Sano, Y.
TITLE
Direct Submission
JOURNAL
Submitted (08-FEB-1997) Yoshitaka Sano, Kyoto Institute of
Technology, Department of Applied Biology; Matsugasaki, Sakyo-ku,
Kyoto 606-8585, Japan (Tel:075-724-7764, Fax:075-724-7764)
2 (sites)
REFERENCE
AUTHORS
Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T. and Kojima, M.
TITLE
Sequences of ten circular ssDNA components associated with the milk
vetch dwarf virus genome
JOURNAL
J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
99094638
MEDLINE
J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
COMMENT
On Oct 29, 1998 this sequence version replaced gi:3798659.
Sequence updated (29-Nov-1997).
FEATURES
source
Location/Qualifiers
1..989
/organism="milk vetch dwarf virus"
/virion
/db_xref="taxon:67585"
/chromosome="5"
/lab_host="Pisum sativum"
6..34
279..285
376..828
/function="unknown"
/feature="putative"
/codon_start=1
/product="viral non-structural protein"
/protein_id="BAA33984.1"
/db_xref="GI:3798660"
/translation="MSVNDYSLDFEATDELVSSEKRLIANDCHDDSDQIVNKKEDLEF
CDMSDKVILKQIFRICYKVRKILDTITLIGCMKYTELKPSSEKSLSILOKKNNIISL
DGNVIGIRMEFFININOLINTCKMTIRIEDVYPICTLIYHANNPVIDI"
stem_loop
TATA_signal
CDS
polyA_signal
BASE COUNT      284 a      167 c      210 g      328 t
ORIGIN

Query Match      66.5%; Score 20.6; DB 14; Length 989;
Best Local Similarity 85.2%; Pred. No. 1.le+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```


OY 5 gcgggggtaataactagcccccgcgcgc 31
 DB 31 GCGGGGTAACTAAGCCCCGCCCT 5

RESULT 7
 LOCUS AB000923/c 990 bp DNA circular VRL 30-OCT-1998
 DEFINITION milk vetch dwarf virus genome segment 4 encoding viral non-structural protein, complete sequence.
 ACCESSION AB000923
 VERSION AB000923.1 GI:3808182
 KEYWORDS viral non-structural protein.
 SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
 ORGANISM Viruses: ssDNA viruses: Nanovirus.
 REFERENCE 1 (bases 1 to 990)
 AUTHORS Sano, Y.
 JOURNAL Direct Submission
 TITLE Submitted (08-FEB-1997) Yoshitaka Sano, Kyoto Institute of Technology, Department of Applied Biology; Matsugasaki, Sakyo-Ku, Kyoto 606-8585, Japan (Tel:075-724-7764, Fax:075-724-7764)
 REFERENCE 2 (siles)
 AUTHORS Sano, Y., Mada, M., Hashimoto, Y., Matsumoto, T., and Kojima, M.
 TITLE Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome
 JOURNAL J. Gen. Virol. 79 (pt 12), 3111-3118 (1998)
 COMMENT On Oct 29, 1998 this sequence version replaced gi:3798657.
 FEATURES
 source
 1..990
 location/Qualifiers
 /organism="milk vetch dwarf virus"
 /villon
 /db_xref="taxon:67585"
 /chromosome="segment 4"
 /lab_host="Pisum sativum"
 6..34
 279..285
 319..928
 /function="unknown"
 /codon_start=1
 /product="viral non-structural protein"
 /protein_id="BA33983.1"
 /db_xref="GI:3798658"
 /translation="MGKRYAHLPLEIREKIVDHOERKKKELEKALDSCRRHEA
 LIIDESPALNSIKFLIALSDYVGNQFNTKRLIRKKDVPKVFGEFDEQHLLY
 GSDMDLSCGELFIPDEDDLTVEGVIVRCSDLQLEFKSLIYIVYVSKHCIMA
 PLSEIVIK"

polyA-signal 820..825
 BASE COUNT 285 a 161 c 226 g 318 t
 ORIGIN

Query Match 66.5%; score 20.6; DB 14; Length 990;
 Best Local Similarity 85.2%; Pred. No. 1.le+03;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 gcgggggtaataactagcccccgcgcgc 31
 DB 31 GCGGGGTAACTAAGCCCCGCCCT 5

RESULT 8
 LOCUS AC098227/c 100542 bp DNA linear HTG 20-DEC-2001
 DEFINITION Rattus norvegicus clone CH230-58J19, *** SEQUENCING IN PROGRESS
 ACCESSION AC098227
 VERSION AC098227.3 GI:17974408
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 100542)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alstbrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbieri, J.,
 Benton, J., Binage, K., Blankenburg, R., Bonin, D., Bouck, J.,
 Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
 Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
 Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
 Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
 Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Gorrell, J.H., Guetara, M., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
 Hernandez, J., Hernandez, O., Hodson, A., Hogues, M., Hollway, C.,
 Hollins, B., Honsi, F., Howard, S., Huber, J., Hulys, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvath, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W.,
 Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mathiney, E., McLeod, M.P., Meador, M.,
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwkw, S.,
 Ogih, M., Okunou, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Prims, E., Pu, L.L.,
 Ruiz, S., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M.,
 Quiles, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N.,
 Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalton, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Wellington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 100542)
 Worley, K.C.
 Direct Submission
 Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:17062637.

COMMENT

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project information
 Center project name: GTVS
 Center clone name: CH230-58J19
 Summary Statistics
 Assembly program: Phrap; version 0.990329first call to
 findphraplist

Consensus quality: 85818 bases at least Q40
 Consensus quality: 94218 bases at least Q30
 Consensus quality: 99376 bases at least Q20
 Estimated insert size: 54170, sum-of-coverage estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 0.6x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a "working draft" sequence. It currently
 * consists of 59 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 4312: contig of 4312 bp in length
 * 4313 4412: gap of unknown length
 * 4413 7638: contig of 3226 bp in length
 * 7639 7738: gap of unknown length
 * 7739 10359: contig of 2621 bp in length
 * 10360 10459: gap of unknown length
 * 10460 13662: contig of 3203 bp in length
 * 13663 13762: gap of unknown length
 * 13763 15839: contig of 2077 bp in length
 * 15840 15939: gap of unknown length
 * 15940 17655: contig of 1716 bp in length
 * 17656 17755: gap of unknown length
 * 17756 19386: contig of 1631 bp in length
 * 19387 19486: gap of unknown length
 * 19487 21390: contig of 1904 bp in length
 * 21391 23084: gap of unknown length
 * 23085 23184: gap of unknown length
 * 23185 25329: contig of 2145 bp in length
 * 25330 25429: gap of unknown length
 * 25430 26940: contig of 1511 bp in length
 * 26941 27040: gap of unknown length
 * 27041 28658: contig of 1618 bp in length
 * 28659 28758: gap of unknown length
 * 28759 30808: contig of 2050 bp in length
 * 30809 32751: contig of 1843 bp in length
 * 32752 32851: gap of unknown length
 * 32852 33882: contig of 1041 bp in length
 * 33883 33992: gap of unknown length
 * 33993 35243: contig of 1251 bp in length
 * 35244 35343: gap of unknown length
 * 35344 36869: contig of 1526 bp in length
 * 36870 36970: gap of unknown length
 * 36971 38969: contig of 2000 bp in length
 * 38970 39059: gap of unknown length
 * 39060 40571: contig of 1502 bp in length
 * 40572 40671: gap of unknown length
 * 40672 41748: contig of 1077 bp in length
 * 41749 41848: gap of unknown length
 * 41849 43764: contig of 1916 bp in length
 * 43765 43864: gap of unknown length
 * 43865 44951: contig of 1087 bp in length
 * 44952 45051: gap of unknown length
 * 45052 47114: contig of 2063 bp in length
 * 47115 47214: gap of unknown length
 * 47215 48401: contig of 1187 bp in length
 * 48402 48501: gap of unknown length
 * 48502 49778: contig of 1277 bp in length
 * 49779 49879: gap of unknown length
 * 49880 51538: contig of 1660 bp in length
 * 51539 51638: gap of unknown length
 * 51639 53492: contig of 1854 bp in length
 * 53493 53592: gap of unknown length
 * 53593 54753: contig of 1161 bp in length
 * 54754 54853: gap of unknown length
 * 54854 56334: contig of 1481 bp in length
 * 56335 56434: gap of unknown length
 * 56435 57728: contig of 1294 bp in length
 * 57729 57828: gap of unknown length
 * 57829 59985: contig of 2167 bp in length
 * 59986 60095: gap of unknown length
 * 60096 61595: contig of 1500 bp in length
 * 61596 63519: contig of 1824 bp in length
 * 63520 63619: gap of unknown length
 * 63620 64853: contig of 1234 bp in length

64854 64953: gap of unknown length
 * 64954 66047: contig of 1094 bp in length
 * 66048 66147: gap of unknown length
 * 66148 67853: contig of 1706 bp in length
 * 67854 67953: gap of unknown length
 * 67954 69623: contig of 1670 bp in length
 * 69624 69723: gap of unknown length
 * 69724 71072: contig of 1349 bp in length
 * 71073 71172: gap of unknown length
 * 71173 72453: contig of 1281 bp in length
 * 72454 72553: gap of unknown length
 * 72554 73633: contig of 1060 bp in length
 * 73634 73733: gap of unknown length
 * 73734 74776: contig of 1043 bp in length
 * 74777 74876: gap of unknown length
 * 74877 76684: contig of 1808 bp in length
 * 76685 76784: gap of unknown length
 * 76785 78395: contig of 1611 bp in length
 * 78396 78495: gap of unknown length
 * 78496 79759: contig of 1264 bp in length
 * 79760 79859: gap of unknown length
 * 79860 80988: contig of 1129 bp in length
 * 80989 81088: gap of unknown length
 * 81089 82146: contig of 1058 bp in length
 * 82147 82247: gap of unknown length
 * 82248 83416: contig of 1170 bp in length
 * 83417 83516: gap of unknown length
 * 83517 84881: contig of 1365 bp in length
 * 84882 84981: gap of unknown length
 * 84982 86401: contig of 1420 bp in length
 * 86402 86501: gap of unknown length
 * 86502 87826: contig of 1325 bp in length
 * 87827 87926: gap of unknown length
 * 87927 89502: contig of 1576 bp in length
 * 89503 89602: gap of unknown length
 * 89603 90937: contig of 1335 bp in length
 * 90938 91037: gap of unknown length
 * 91038 92578: contig of 1541 bp in length
 * 92579 92678: gap of unknown length
 * 92679 94135: gap of unknown length
 * 94136 94234: contig of 1456 bp in length
 * 94235 95305: contig of 1071 bp in length
 * 95306 95405: gap of unknown length
 * 95406 97064: contig of 1659 bp in length
 * 97065 97164: gap of unknown length

Query Match 65.8%; Score 20.4; DB 2; Length 100542;
 Best Local Similarity 80.0%; Pred. No. 4.6e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 gccgcgggggtatctactagcccgccgct 31
 DB 75204 gccgcgggggtatctagccgcgcgcgct 75175

RESULT 9

AC098923 179030 bp DNA linear HTG 20-DEC-2001
 LOCUS Rattus norvegicus clone CH230-154B15, *** SEQUENCING IN PROGRESS
 DEFINITION *** 68 unordered pieces.

AC098923
 AC098923.4 GI:17973852
 VERSION HTG: HTGS, PHASEL.
 KEYWORDS Norway rat.
 SOURCE Rattus norvegicus
 ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 179030)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alstrooms, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J.,
 Benton, J., Blum, K., Blankensbury, K., Bonnin, D., Bouck, J.,
 Bowie, S., Brieval, M., Brown, E., Brown, M., Bryant, N.P., Buhan, C.,

Burch, P., Burdett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, Z., Chowdhury, I., Christopoulos, C., Cleeland, C.D., Cox, C., Coyle, M.D., Dabholme, S.R., David, R., Davila, M.L., Davis, C., Day-Carroll, L., Dedich, D.A., Delaney, K.J., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferriguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guervara, W., Gunnatane, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Huijks, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Koryak, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Lea, B., Lewis, L.C., Lewis, L., Li, Z., Licharge, O., Lieu, C., Liu, C., Liu, J., Liu, W., Louieged, H., Lozado, R.D., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mappa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, A., Morris, S., Moser, M., Neal, D., Newton, J., Nextson, N., Nguyen, A., Nguyen, N., Nguyen, N., Nicolson, E., Nwokewo, S., Oguh, M., Okunolu, G., Oigunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojes, A., Rojibockan, I., Rolfe, M., Ruiz, S., Severy, G., Scherer, S., Scott, G., Shen, H., Shoochari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Sytek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Tellro, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 179030)

Worley, K.C.

Direct Submission
Submitted (06-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17064677.

Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: CH330-154B15
Center clone name: CH330-154B15

Summary Statistics

Assembly Program: Phrap; version 0.990329First call to
IndelPhraplast

Consensus quality: 137096 bases at least Q40
Consensus quality: 145080 bases at least Q30
Consensus quality: 152950 bases at least Q20
Estimated insert size: 138980; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agrote-tp estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 7467: contig of 7467 bp in length

Wed Jun 26 08:07:42 2002

us-09-462-955b-2.rge

Page 10

BASE COUNT 144 a 126 c 129 g 120 t
ORIGIN

Query Match 63.9%; Score 19.8; DB 5; Length 519;
Best Local Similarity 91.3%; Pred. No. 2.6e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3 ccgcgggggtataactagcccc 25
||||||| |||||
Db 288 CCGCGGGGAAAGACTAGCCCC 310

Search completed: June 25, 2002, 15:39:46
Job time: 3607 sec